

# Análise exploratória de dados

## Parte 4

Prof.: Eduardo Vargas Ferreira



---

# Gráficos para variáveis quantitativas

# Gráficos para variáveis quantitativas

**Histograma**



**Gráfico de linhas**

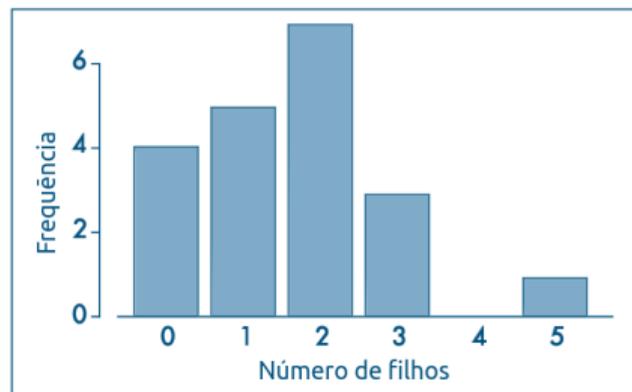


# Gráfico de barras

- ▶ O gráfico de barras também pode ser usado em variáveis quantitativas (neste caso discreta).

Nº de filhos	Frequência
$x_i$	$n_i$
0	4
1	5
2	7
3	3
5	1
Total	20

**Figura:** Gráfico de barras para a variável: número de filhos.



Fonte: Estatística Básica (Bussab e Morettin, 2017)

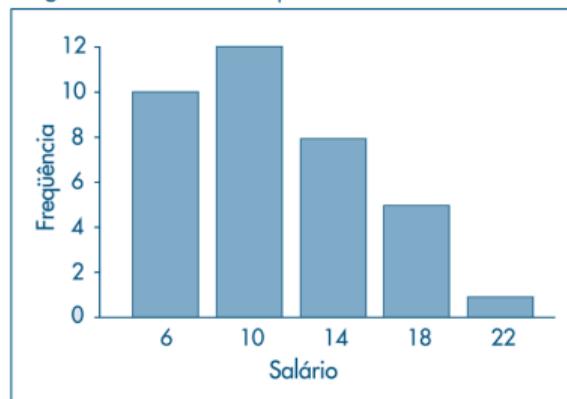
- ▶ A ordem das barras é relevante, já que a variável é ordinal.

# Gráfico de barras

- ▶ No caso de variáveis contínuas, também podemos representá-las através de barras.

Classe de salários	Ponto médio	Frequência
4,00 ┆ 8,00	6	10
8,00 ┆ 12,00	10	12
12,00 ┆ 16,00	14	8
16,00 ┆ 20,00	18	5
20,00 ┆ 24,00	22	1
Total	-	36

Figura: Gráfico em barras para a variável salários.



Fonte: Estatística Básica (Bussab e Morettin, 2017)

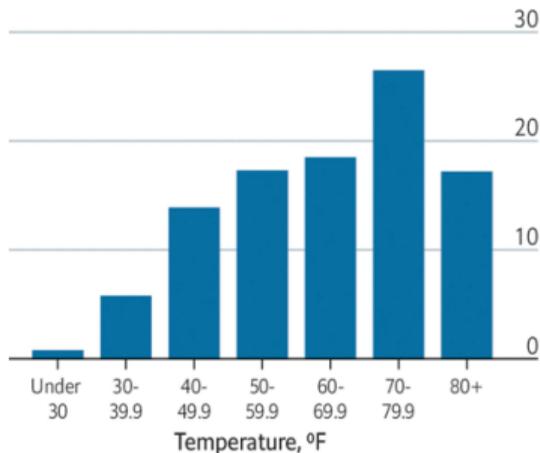
- ▶ Tal artifício faz com que se perca informações relevantes. Uma alternativa neste caso é o **histograma**.

# Exemplo: calor aumenta risco de violência nas prisões americanas

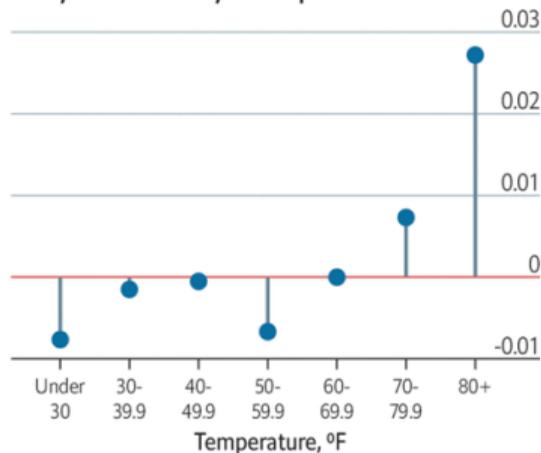
## Hot and homicidal

Mississippi Department of Corrections, 2004-2010

Share of days by temperature, %



Modelled change in number of violent acts for every additional day of temperature at:



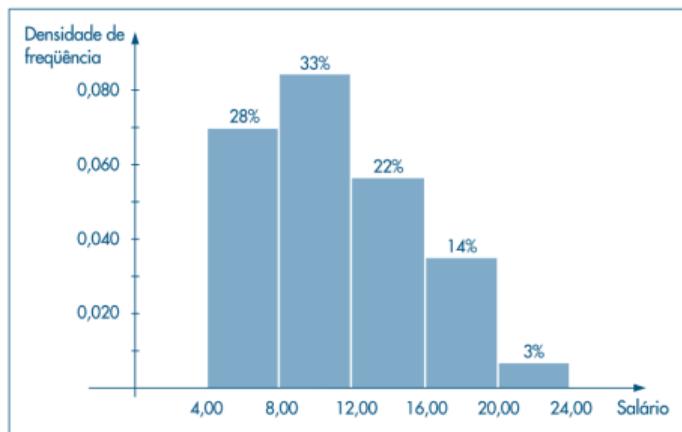
Source: "The Causal Effect of Heat on Violence", by Anita Mukherjee & Nicholas Sanders, NBER Working Paper, July 2021

The Economist

# Histograma

- ▶ Trata-se de um gráfico de barras contíguas, com bases proporcionais aos intervalos das classes, e a área dos retângulos proporcional à frequência.

Classe de salários	Porcentagem
4,00 ┆ 8,00	27.78
8,00 ┆ 12,00	33.33
12,00 ┆ 16,00	22.22
16,00 ┆ 20,00	13.89
20,00 ┆ 24,00	2.78
Total	100



Fonte: Estatística Básica (Bussab e Morettin, 2017)

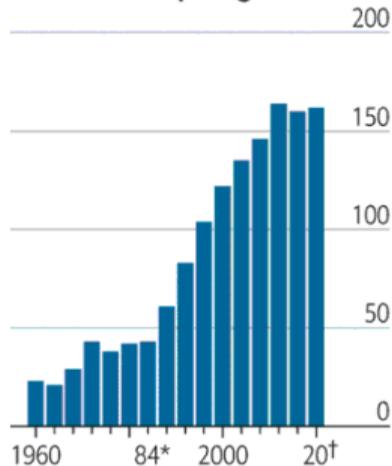
- ▶ Pode-se usar tanto a frequência absoluta,  $n_i$ , como a relativa,  $f_i$ .

# Exemplo: os jogos paralímpicos estão cada vez mais fortes

## Olympian heights

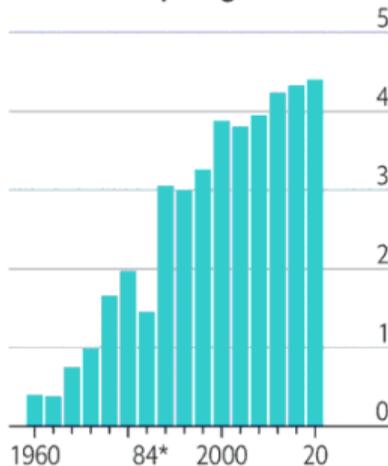
The Paralympics

Countries competing



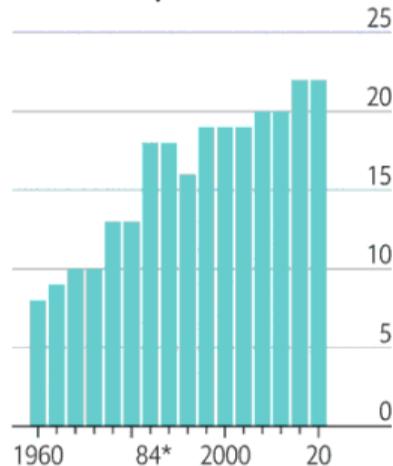
Source: International Paralympic Committee

Athletes competing, '000



\*Average of the two games held that year

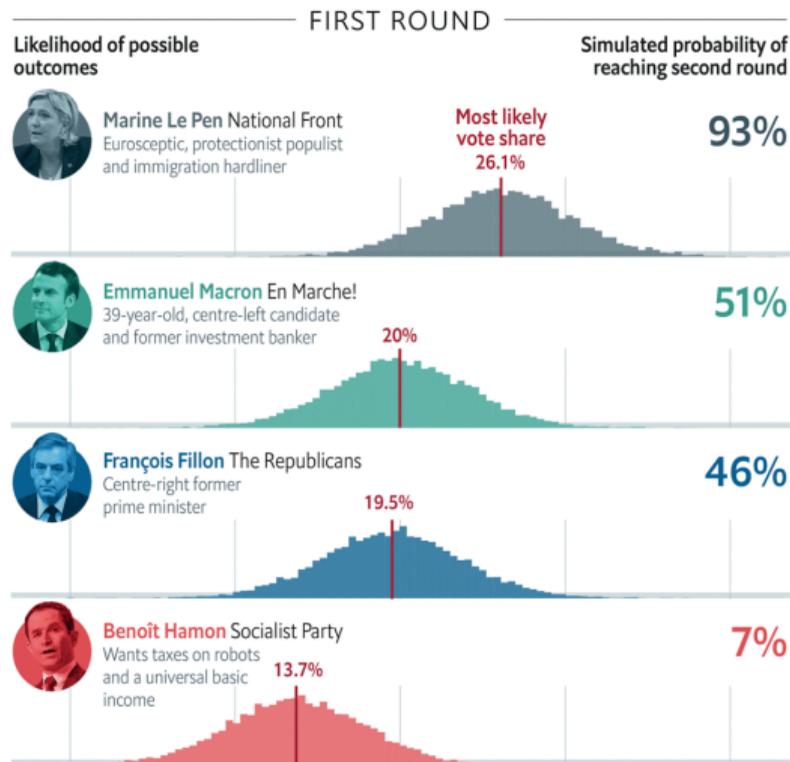
Number of sports



†Includes a team of refugee athletes

economist.com

# Exemplo: eleição presidencial da França



# Histograma

- ▶ Esse histograma alisado é útil para ilustrar rapidamente qual o tipo de comportamento que se espera para a distribuição de uma dada variável.

Figura: Histograma da variável: salários.

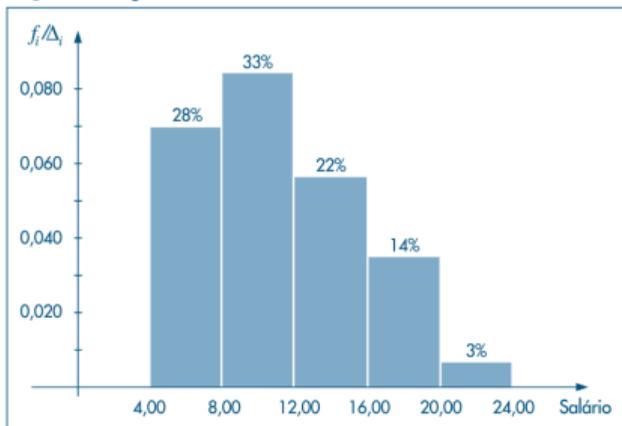
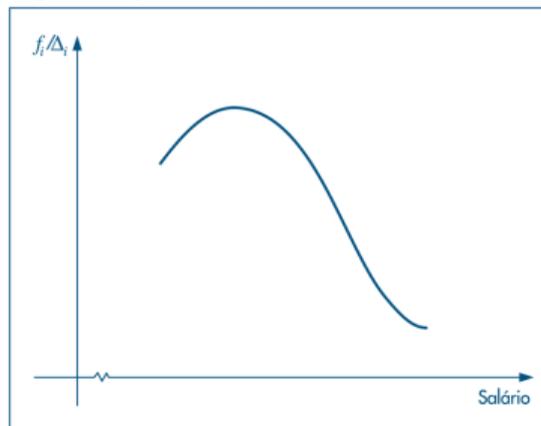


Figura: Histograma alisado para a variável: salário.



Fonte: Estatística Básica (Bussab e Morettin, 2017)

- ▶ A interpretação é a mesma do histograma. Assim, nas regiões onde a curva é mais alta, significa uma maior densidade de observações.

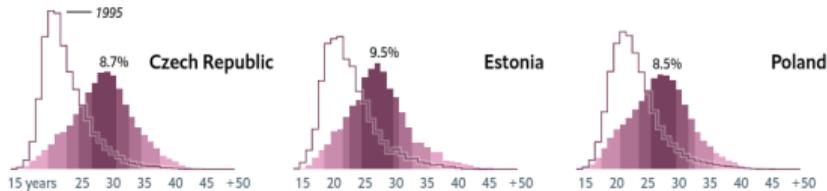
# Exemplo: idade das mulheres no primeiro parto

## Thirty is the new twenty

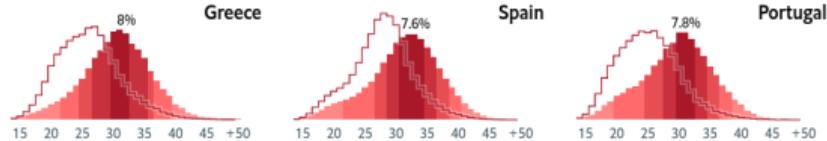
Ages of women at first birth, selected countries



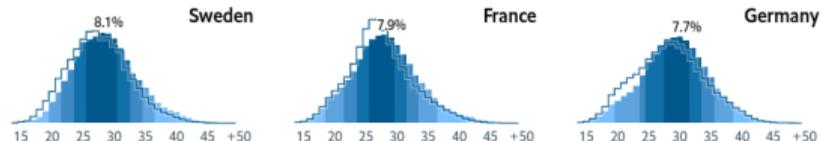
### Eastern Europe



### Southern Europe



### Western Europe



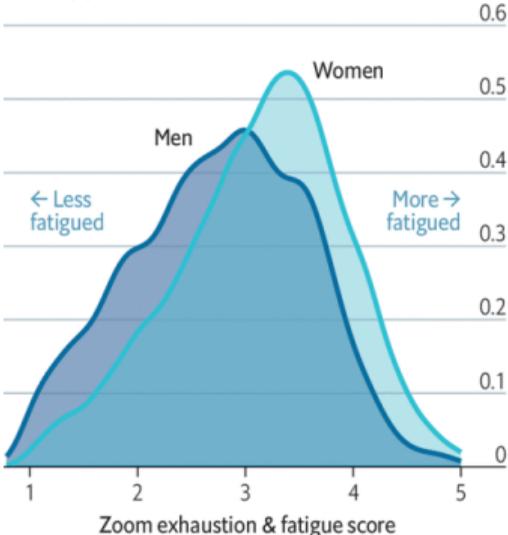
economist.com

# Exemplo: fadiga devido à muitas reuniões

## Zoom and gloom

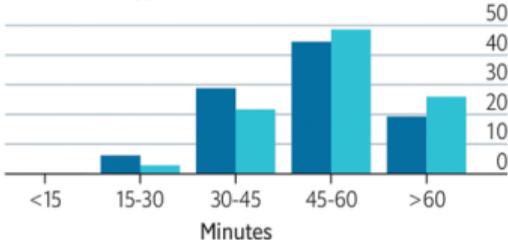
Videoconferencing habits by sex\*

Density plot

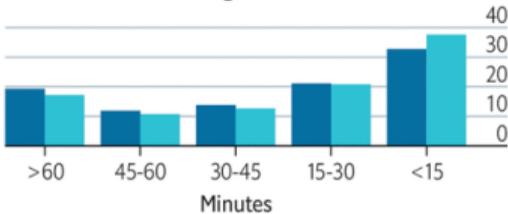


Source: "Nonverbal mechanisms predict zoom fatigue and explain why women experience higher levels than men", by G. Fauville, M. Luo, A. Queiroz, J. Bailenson, & J. Hancock

Duration of typical video call, %



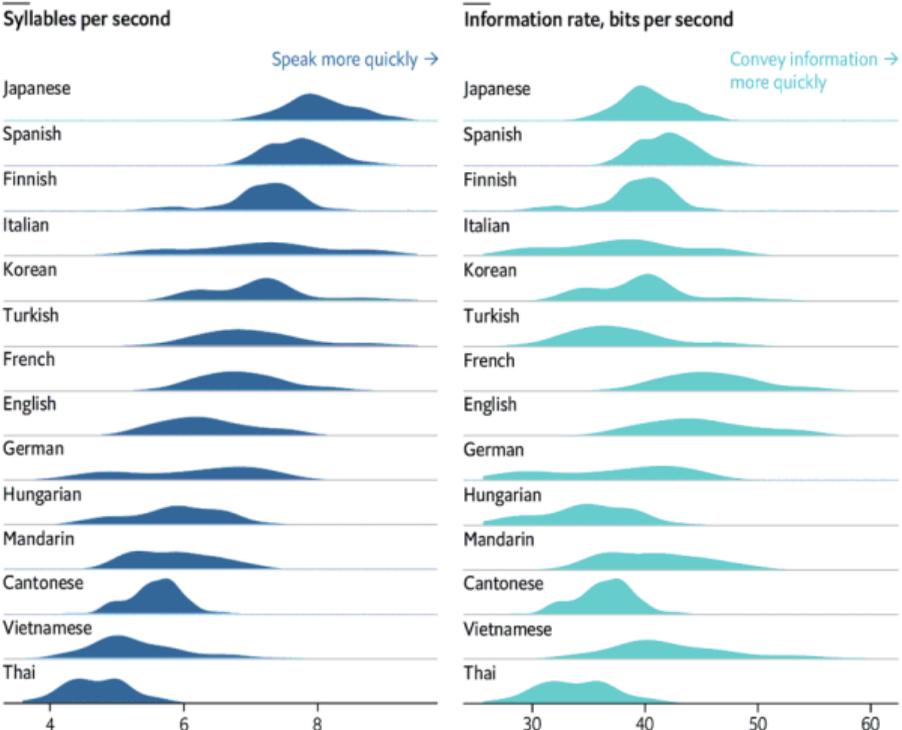
Time between meetings, %



\*Survey of 10,332 people conducted February 22nd to March 12th 2021

The Economist

# Exemplo: alguns idiomas são falados mais rápido que outros?



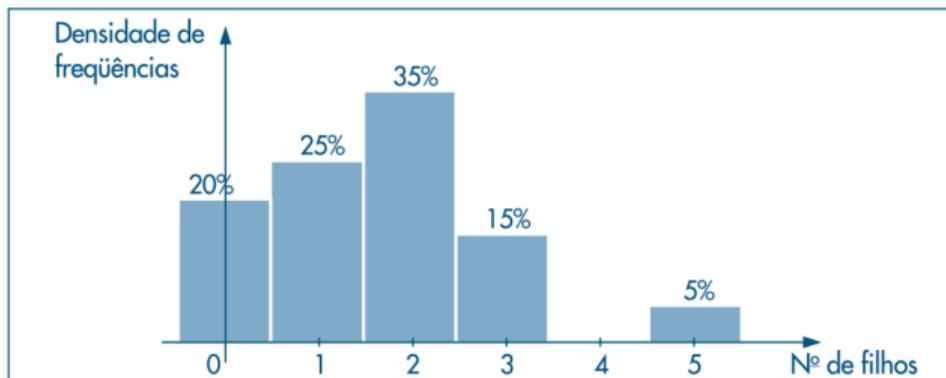
Source: "Different languages, similar encoding efficiency: Comparable information rates across the human communicative niche" by Christophe Coupé, Yoon Mi Oh, Dan Dediu and François Pellegrino, *Science Advances* (2019)

economist.com

# Histograma

- ▶ Alternativamente podemos construir um histograma para variáveis discretas.

**Figura:** Histograma da variável: número de filhos.



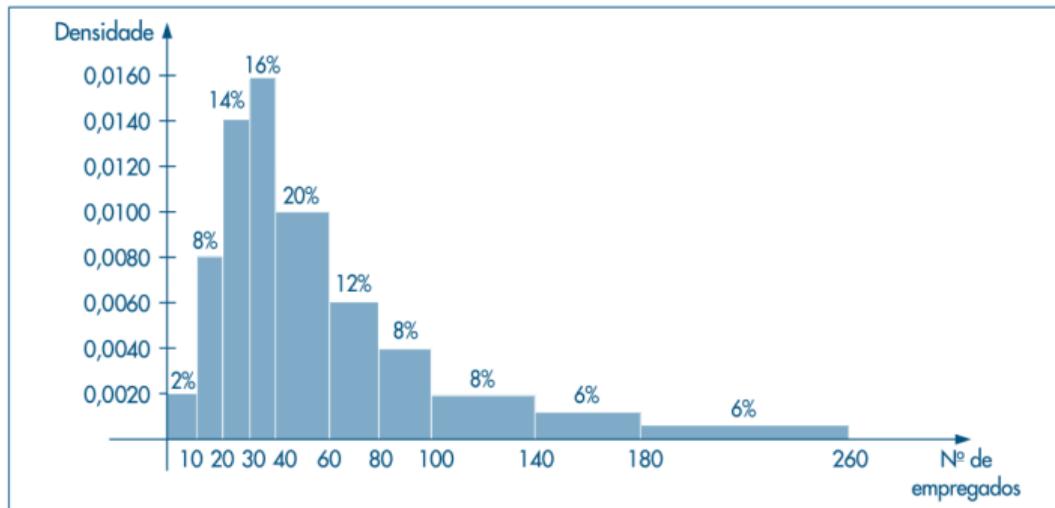
Fonte: Estatística Básica (Bussab e Morettin, 2017)

- ▶ O gráfico é suficientemente auto-explicativo, de modo que omitiremos detalhes sobre sua construção.

# Histograma

- ▶ É muito comum o uso de classes com tamanhos desiguais no agrupamento.

**Figura:** Histograma de 250 empresas segundo o número de funcionários



Fonte: Estatística Básica (Bussab e Morettin, 2017)

- ▶ Tal abordagem torna o gráfico mais informativo, especialmente nas regiões de baixa frequência.

# Gráficos para variáveis quantitativas

**Histograma**

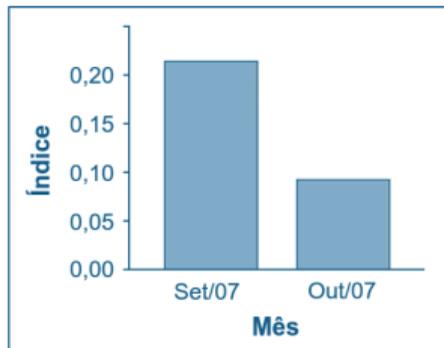


**Gráfico de linhas**



## IPC-Fipe recua para 0,08% em outubro

O Índice de Preços ao Consumidor (IPC) da Fundação Instituto de Pesquisas Econômicas (Fipe), da USP, fechou o mês de outubro com variação de 0,08% na cidade de São Paulo. O índice apresentou significativo recuo ante a taxa setembro (0,24%) e ficou abaixo das expectativas dos analistas consultados pela Agência Estado, que iam de 0,11% a 0,16%.

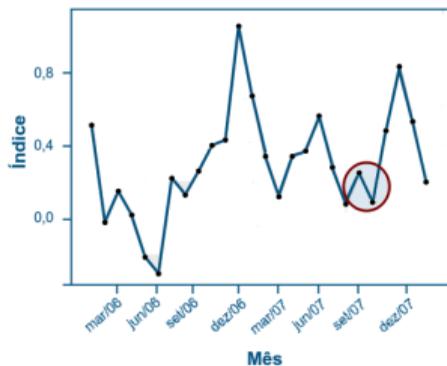


Fonte: Estatística Básica (Bussab e Morettin, 2017)

- ▶ A inflação está caindo?
- ▶ As variações são grandes ou pequenas?
- ▶ Em relação a quê?
- ▶ O que esperar para o mês seguinte?

## IPC-Fipe recua para 0,08% em outubro

O Índice de Preços ao Consumidor (IPC) da Fundação Instituto de Pesquisas Econômicas (Fipe), da USP, fechou o mês de outubro com variação de 0,08% na cidade de São Paulo. O índice apresentou significativo recuo ante a taxa setembro (0,24%) e ficou abaixo das expectativas dos analistas consultados pela Agência Estado, que iam de 0,11% a 0,16%.

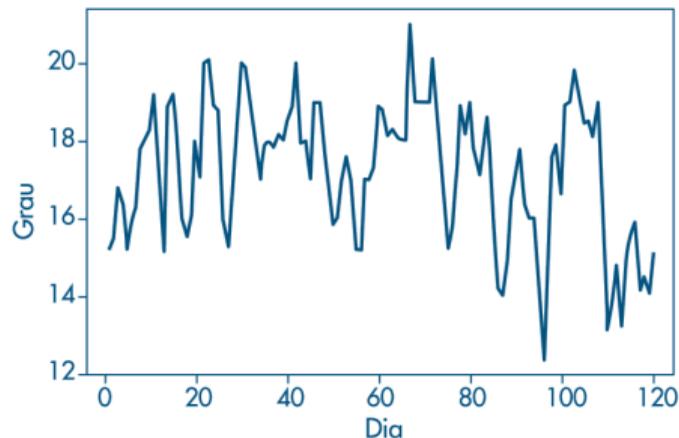


- ▶ A inflação está caindo?
- ▶ As variações são grandes ou pequenas?
- ▶ Em relação a quê?
- ▶ O que esperar para o mês seguinte?

# Gráfico de Linhas

- ▶ Permite representar séries longas, auxiliando a detectar tanto flutuações quanto analisar tendências;

**Figura:** Dados de temperatura de São Paulo.



Fonte: Estatística Básica (Bussab e Morettin, 2017)

- ▶ Sua aplicação é mais indicada para representações de séries temporais, sendo também conhecidos como gráficos de séries cronológicas;

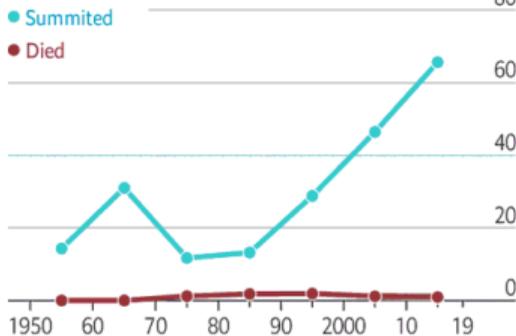
# Exemplo: mais pessoas estão escalando o Monte Everest

## Getting better all the time

Mount Everest, first-time summit attempts

### Summit and death rates, %

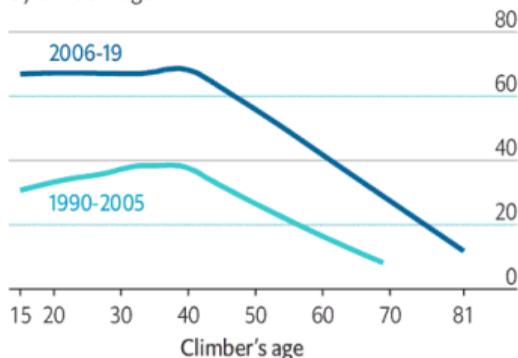
By decade



Source: "Mountaineers on Mount Everest: effects of age, sex, experience, and crowding on rates of success and death" by Huey et al., PLoS ONE 2020

### Successful summits, %

By climber's age



economist.com

# Gráfico de frequência acumulada

- ▶ Uma medida usada para descrever dados quantitativos é a frequência acumulada, que indica quantos elementos estão abaixo de um certo valor.

Nº de filhos	Frequência	Frequência acumulada
0	4	4
1	5	9
2	7	16
3	3	19
5	1	20

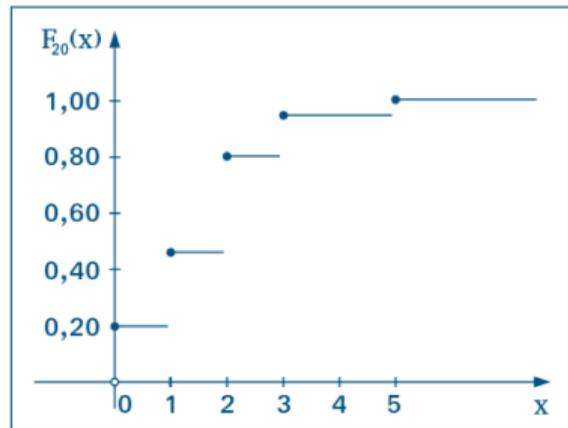
$$F_{20}(x) = \begin{cases} 0,00, & \text{se } x < 0 \\ 0,20, & \text{se } 0 \leq x < 1 \\ 0,45, & \text{se } 1 \leq x < 2 \\ 0,80, & \text{se } 2 \leq x < 3 \\ 0,95, & \text{se } 3 \leq x < 5 \\ 1,00, & \text{se } x \geq 5 \end{cases}$$

- ▶ Este gráfico pode ser usado para fornecer informações adicionais. P. ex., qual o salário  $s$  tal que 50% dos funcionários ganhem menos do que  $s$ .

# Gráfico de frequência acumulada

- ▶ Uma medida usada para descrever dados quantitativos é a frequência acumulada, que indica quantos elementos estão abaixo de um certo valor.

$$F_{20}(x) = \begin{cases} 0,00, & \text{se } x < 0 \\ 0,20, & \text{se } 0 \leq x < 1 \\ 0,45, & \text{se } 1 \leq x < 2 \\ 0,80, & \text{se } 2 \leq x < 3 \\ 0,95, & \text{se } 3 \leq x < 5 \\ 1,00, & \text{se } x \geq 5 \end{cases}$$



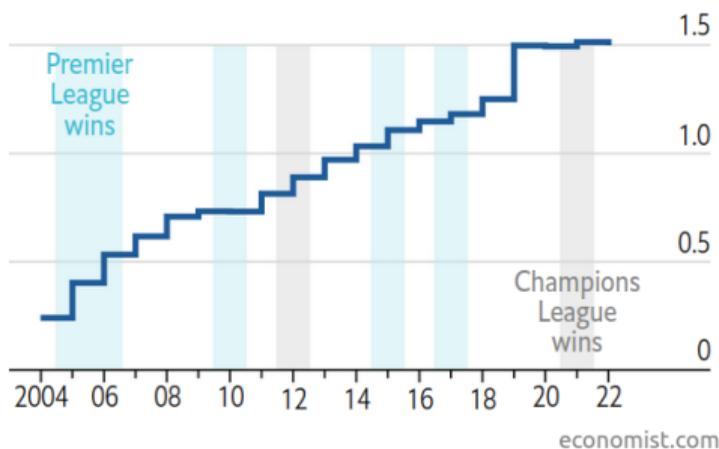
- ▶ Este gráfico pode ser usado para fornecer informações adicionais. P. ex., qual o salário  $s$  tal que 50% dos funcionários ganhem menos do que  $s$ .

# Exemplo: investimento acumulado no Chelsea durante os anos

## A costly prize

English Premier League, football club finances

Roman Abramovich's cumulative investment in Chelsea, £bn

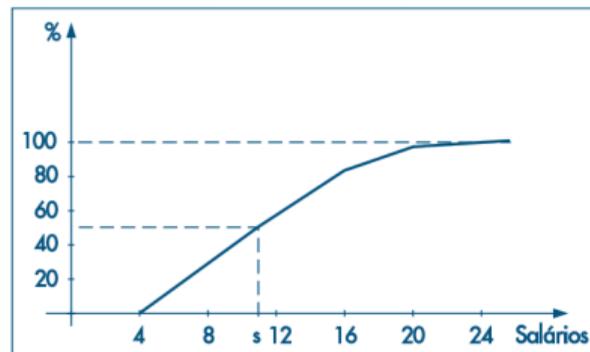


# Gráfico de frequência acumulada

- ▶ Esta definição também vale para variáveis quantitativas contínuas. Neste caso, consideramos o exemplo da variável salário.

Classe de salários	Frequência acumulada	Porcentagem acumulada
4,00 ┆ 8,00	10	27,78
8,00 ┆ 12,00	22	61,11
12,00 ┆ 16,00	30	83,33
16,00 ┆ 20,00	35	97,22
20,00 ┆ 24,00	36	100,00

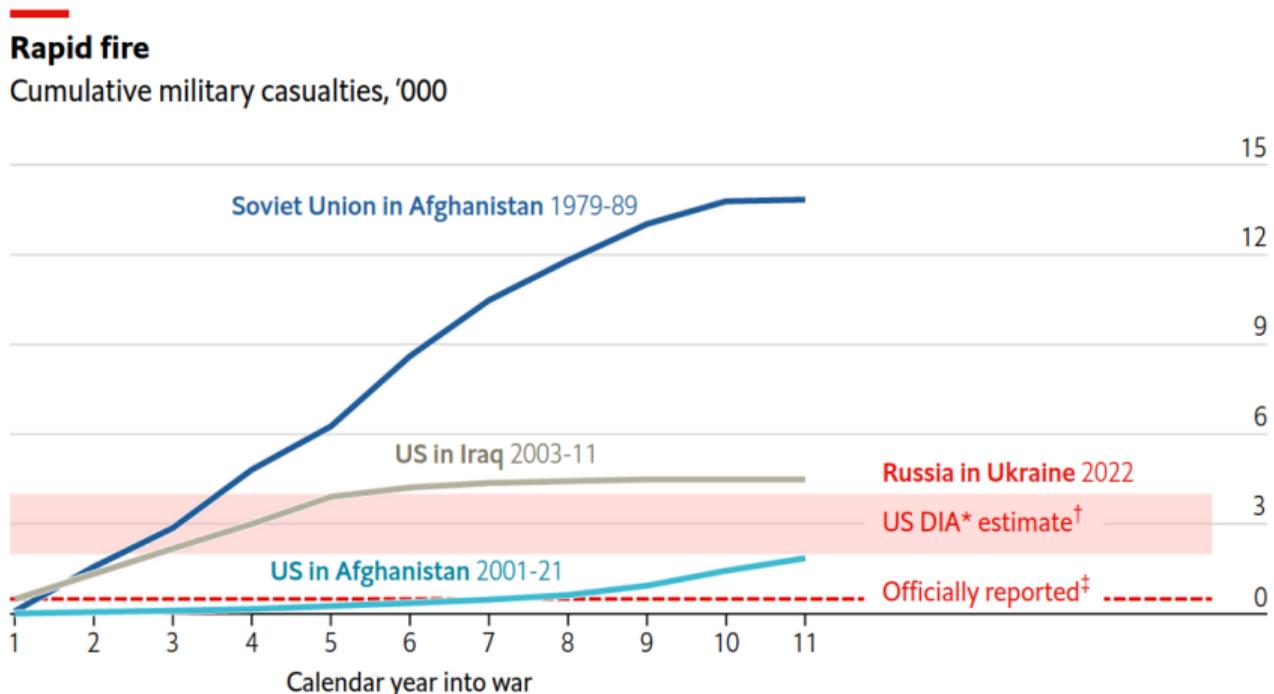
Figura: Porcentagens acumuladas para o salário.



Fonte: Estatística Básica (Bussab e Morettin, 2017)

- ▶ Este gráfico pode ser usado para fornecer informações adicionais. P. ex., qual o salário  $s$  tal que 50% dos funcionários ganhem menos do que  $s$ .

# Exemplo: mortes de soldados em guerras



\*Defence Intelligence Agency †First 11 days of war  
‡First 7 days of war

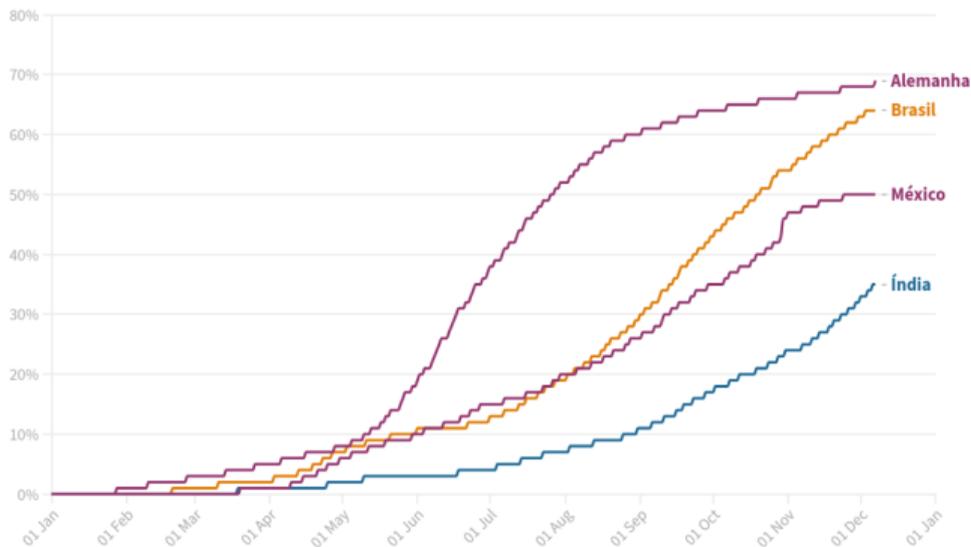
Sources: Brookings Institution; Scribd  
economist.com

# Exemplo: percentual de vacinados

% da população totalmente vacinada

Brasil × Índia × Alemanha × México × Digite um país

Brasil Alemanha Índia México



gazetadopovo.com.br

# Referências

- ▶ Bussab, WO; Morettin, PA. Estatística Básica. São Paulo: Editora Saraiva, 2006 (5ª Edição).
- ▶ Magalhães, MN; Lima, ACP. Noções de Probabilidade e Estatística. São Paulo: EDUSP, 2008.

